

### **REMARKS/ARGUMENTS**

These remarks are made in response to the Office Action of March 25, 2008 (Office Action). As this response is timely filed within the 3-month shortened statutory period, no fee is believed due. However, the Examiner is expressly authorized to charge any deficiencies to Deposit Account No. 50-0951.

### **Claim Rejections – 35 USC § 103**

In the Office Action, Claims 1-3, 8-10, 13-15, 20-22, 25, and 26 were rejected under 35 U.S.C. § 103(a) as being unpatentable over U.S. Patent 5,864,814 to Yamazaki (hereinafter Yamazaki) in view of U.S. Patent 5,842,167 to Miyatake, *et al.* (hereinafter Miyatake), and further in view of U.S. Patent 6,366,883 to Campbell, *et al.* (hereinafter Campbell) and U.S. Patent 7,043,433 to Hejna, Jr. (hereinafter Hejna).

Applicants respectfully disagree with the rejections and thus have not amended the claims.

### **Aspects of Applicants' Invention**

It may be helpful to reiterate certain aspects of Applicants' invention prior to addressing the cited references. One embodiment of the invention, as typified by amended Claim 1, is a computer-implemented method for debugging and tuning synthesized audio.

The method can include the steps of (a) receiving a user-supplied text with a visual user interface; (b) generating synthesized audio generated from concatenated phonetic units, the synthesized audio being a voice rendering of the user-supplied text; (c) displaying a waveform corresponding to the synthesized audio generated from concatenated phonetic units; (d) displaying parameters corresponding to at least one of

the phonetic units, the parameters including configuration parameters comprising at least one weight for adjusting at least one search cost function, and the at least one weight comprising at least one of a pitch cost weight and a duration cost weight; (e) displaying an original recording containing a selected phonetic unit; (f) receiving an editing input from the user; (g) adjusting at least one configuration parameter in accordance with the editing input and storing the at least one configuration parameter in a text-to-speech engine configuration file, wherein adjusting includes repositioning a phonetic alignment marker; (h) highlighting in the display of the original recording at least one user-selected phonetic unit; (i) correcting elements of a text-to-speech segment dataset of parameters corresponding to a segment of the synthesized audio identified as be problematic; (j) generating a new synthesized waveform corresponding to one or more adjusted parameters; and (k) repeating steps (b)-(j) until a desired synthesized output is generated. See, e.g., Specification, paragraphs [0031] to [0034]; see also Figs. 2 and 4

### **The Claims Define Over The Prior Art**

The present invention disclosed a method and a system for identifying and correcting sources of problems in synthesized speech, which is generated using a concatenative text-to-speech (CTTS) technique. In particular, the present invention provides modules and tools which can be used to quickly identify problem audio segments and edit parameters associated with the audio segments. For example, such problem identification and parameter editing can be performed using a graphical user interface (GUI). In particular, voice configuration files containing general voice parameters and text-to-speech (TTS) segment datasets having parameters associated with the problem audio segments can be automatically presented within the GUI for editing. In comparison to traditional methods of identifying and correcting synthesized audio

segments, the method of the present invention is much more efficient and less tedious. See specification, paragraph [0017].

Yamazaki concerns a technology in which a voice is divided into voice source information and voice route information (voice tone information) corresponding to the voice source information to facilitate transfer over the Internet. The voice source information and voice tone information corresponding to each other are then synthesized into a voice when desired at the client end. The object of Yamazaki is to obtain an optimal correspondence between the voice-generating information and the voice tone information. Therefore, the subject matter of Yamazaki has nothing to do with the object of the present invention, namely identifying and correcting bad audio segments in a synthesized speech by editing parameters using a GUI.

Fig. 26-33 of Yamazaki are views each showing the state shift of an operation screen in the processing for making new voice-generating information. During the process as shown in Fig. 26-33, the pitch and velocity of the phonemes can be adjusted to reproduce a synthesized voice with a waveform as close to the waveform of the original voice as possible. Therefore, Yamazaki concerns reproducing a synthesized voice that is as similar to the original voice as possible. In contrast, the present invention concerns debugging and tuning of synthesized speech without comparing to the original speech because an original speech does not exist. It is noted that in Yamazaki the synthesized voice is synthesized from voice source information and voice tone information received via Internet, whereas in the present invention the synthesized speech is synthesized from phonetic units in a phonetic data store. Therefore, the term "synthesize" as used in Yamazaki is not used in the same sense as it pertains to the claimed features of the present invention.

It is also noted that in the present invention, the changes to the segment parameters will be saved in the phonetic data store and can be used for future synthesis. Thus, after

certain repetitions of debugging and tuning, the quality of the phonetic data in the data store improves. Eventually the debugging and tuning need not be continued. In contrast, in Yamazaki the process of making new voice-generating information has to be repeated each time new voice source information and voice tone information are received from the Internet.

The other cited references do not make up for the deficiencies of Yamazaki as discussed above. Accordingly, Applicants believe that the cited references, alone or in combination, fail to disclose or suggest the concept of the present invention, namely generating synthesized audio using concatenated phonetic units from a user-supplied text received in a visual user interface and tuning the synthesized audio by adjusting parameters (including parameters of the phonetic units and text-to-speech engine configuration parameters) displayed in the visual user interface until desired synthesized output is generated, as recited in independent Claims 1, 13, and 25. Applicants therefore respectfully submit that independent Claims 1, 13, and 25 define over the prior art. Furthermore, as each of the remaining claims depends from Claim 1, 13, or 25 while reciting additional features, Applicants further respectfully submit that the remaining claims likewise define over the prior art.

Applicants thus respectfully request that the claim rejections under 35 U.S.C. § 103 be withdrawn.

### CONCLUSION

Applicants believe that this application is now in full condition for allowance, which action is respectfully requested. Applicants request that the Examiner call the undersigned if clarification is needed on any matter within this Amendment, or if the

Examiner believes a telephone interview would expedite the prosecution of the subject application to completion.

Respectfully submitted,

AKERMAN SENTERFITT

Date: June 25, 2008

/Richard A. Hinson/  
Gregory A. Nelson, Registration No. 30,577  
Richard A. Hinson, Registration No. 47,652  
Yonghong Chen, Registration No. 56,150  
Customer No. 40987  
Post Office Box 3188  
West Palm Beach, FL 33402-3188  
Telephone: (561) 653-5000